



next

NATIONAL EXPERIMENTAL  
THERAPY PARTNERSHIP

# **NEXT Bioinformatics**

**Activity Report: 2015 - 2018**

February 2019  
Version 1



## Hospitals



Rigshospitalet



Herlev  
Hospital



Region of  
Southern Denmark  
OUH  
Odense  
University Hospital



Aarhus University Hospital



AALBORG UNIVERSITY HOSPITAL

## Universities

UNIVERSITY OF  
COPENHAGEN



UNIVERSITY OF  
SOUTHERN DENMARK



AARHUS UNIVERSITY



AALBORG UNIVERSITY  
DENMARK



# Table of Contents

<b>1. INTRODUCTION</b>	<b>4</b>
BACKGROUND OF NEXT BIOINFORMATICS . . . . .	4
WHAT PROBLEMS DID NEXT BIOINFORMATICS TRY TO SOLVE? . . . . .	5
WHAT SOLUTIONS DID NEXT BIOINFORMATICS ARRIVE AT? . . . . .	5
<b>2. NETWORK ACTIVITIES</b>	<b>7</b>
WORKSHOPS . . . . .	7
<b>3. COMMON STANDARDS FOR BIOINFORMATICS WORKFLOWS</b>	<b>8</b>
NEXT COMMON STANDARDS . . . . .	8
IMPLEMENTATION . . . . .	9
FUTURE DIRECTIONS . . . . .	10
<b>4. BIOMARKER STRATIFICATION OF PATIENTS</b>	<b>11</b>
SUB-CLASSIFICATION OF CANCERS BY MACHINE LEARNING METHODS	11
SYSTEMS BIOLOGY AND FUNCTIONAL STUDIES . . . . .	12
IMPLEMENTATION FOR CLINICAL USE . . . . .	12
<b>5. VARIANT DATABASE</b>	<b>14</b>
PLATFORM . . . . .	14
DATA COLLECTION . . . . .	14
VARIANT SEARCH . . . . .	18
PRECISION CANCER MEDICINE REPORT . . . . .	19
<b>6. DISCUSSION</b>	<b>20</b>
MAIN RESULTS OF NEXT BIOINFORMATICS . . . . .	20
CRITICAL ASSESSMENT OF THE WORK . . . . .	21
COMPARISON WITH OTHER INITIATIVES . . . . .	21
CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK . . . . .	21
<b>7. PUBLICATIONS</b>	<b>23</b>
<b>8. ACKNOWLEDGEMENTS</b>	<b>25</b>
<b>9. PARTICIPANTS</b>	<b>26</b>
RIGSHOSPITALET/COPENHAGEN UNIVERSITY . . . . .	26
HERLEV UNIVERSITY HOSPITAL/COPENHAGEN UNIVERSITY . . . . .	26
ODENSE UNIVERSITY HOSPITAL/UNIVERSITY OF SOUTHERN DENMARK	26
AARHUS UNIVERSITY HOSPITAL/AARHUS UNIVERSITY . . . . .	26
AALBORG UNIVERSITY HOSPITAL/AALBORG UNIVERSITY . . . . .	26
EXTERNAL DELEGATES . . . . .	27



# 1. Introduction

---

## **Background of NEXT Bioinformatics**

National Experimental Therapy Partnership (NEXT) is a public-private partnership in clinical research comprising the regions of Denmark, Danish universities, 12 pharmaceutical companies, and one GTS institute. In addition, NEXT has several associated partners. The partnership was established on November 1, 2014, as an INNO+ partnership. Innovation Fund Denmark invested DKK 50 million in the partnership over a five-year period, and the partners invested DKK 114 million. NEXT's objective is to make Denmark the pharmaceutical industry's first choice for early clinical trials of new drugs for patients, with a focus on proof-of-concept trials. NEXT aims to optimize all processes from the start-up to close-out of clinical trials, and the optimization of legal and regulatory processes is highly prioritized.

The detection of cancer that is refractory to standard therapy using molecular biomarkers and the selection of candidates for early-phase clinical trials is becoming more and more important. Many of these methods are based on small gene panels that vary from one cancer type to another. These genomic analyses are not cost effective and it is difficult to systematize achieved knowledge. NEXT Bioinformatics was established as a part of NEXT to optimize the process of extensive genomic profiling (whole exome (DNA) and transcriptome (RNA) sequencing) for drug resistance detection and to make Denmark attractive for clinical trials. NEXT invested DKK 5 million in this initiative and the partners invested an additional DKK 5 million.

NEXT Bioinformatics is based at the Cancer Data Science Laboratory of Aalborg University Hospital (Aalborg UH) and Aalborg University (AAU), and is coordinated by Professor Martin Bøgsted. NEXT Bioinformatics also consists of the Center for Genomic Medicine at Rigshospitalet (RH), the Department of Clinical Genetics, Odense University Hospital (OUH), the Department of



Molecular Medicine, Aarhus University Hospital (AUH), and the Department of Pathology of Herlev University Hospital (HUH).

NEXT Bioinformatics has established a network of leading bioinformatics experts in Denmark who focus on precision medicine in oncology and haematology by generating, validating, and implementing common standards for patient classification based on genome analyses. NEXT Bioinformatics is internationally competitive and an attractive collaboration partner for companies due to its close affiliation with the clinic.

## **What problems did NEXT Bioinformatics try to solve?**

Before NEXT Bioinformatics was founded, bioinformaticians worked in small groups at the university hospitals in Denmark on their local bioinformatics workflows, computer systems, and data storage solutions, leading to inhomogeneous data storage across the nation, a lack of overview and a lack of a direct link between bioinformatics analyses and clinical users. The need to establish a network of bioinformaticians dedicated to harmonizing the work was evident. Moreover, there was a need to aid biomarker discovery of drug resistance based on material from Danish biobanks. Finally, as many biomarker detections are not stored in searchable databases, there is a barrier to their usage in trial design and recruitment. To solve these issues, the strategy of NEXT Bioinformatics was to harmonize bioinformatics workflows across Denmark, cross-fertilize biomarker development, and establish a searchable database of detected variants.

## **What solutions did NEXT Bioinformatics arrive at?**

A start-up meeting was held March 2015 in Aalborg, where the network of Danish clinical bioinformaticians was created and coined NEXT Bioinformatics. The group immediately began to develop common guidelines for bioinformatics workflows. Subsequently, there were nine workshops at which the status of the work was discussed and the developed biomarkers were presented for mutual inspiration.



---

During the project, a variant database based on the findable, accessible, interoperable, and reusable (FAIR) principles was established, which is now in use in the North Denmark Region and is underway in the rest of the country. The work of NEXT Bioinformatics has had great impact on clinical bioinformatics in Denmark, as we now rely on harmonized workflows and have outlined searchable databases. The knowledge in the database is now utilized at Aalborg UH and is being used as inspiration for development of databases in the Danish National Genome Center.



---

## 2. Network activities

---

### Workshops

Half-yearly workshops held during the project period:

**Start-up meeting: NEXT Interdisciplinary Bioinformatics Activity meeting**

March 3, 2015, Aalborg University Hospital, Aalborg

**Workshop #1: Classification and in vitro models**

August 26, 2015, Odense University Hospital, Odense

**Workshop #2: CUP Classifiers**

September 30, 2015, Aarhus University Hospital, Aarhus

**Workshop #3: DNA-based technologies in personalized medicine and data management**

January 26, 2016, Rigshospitalet, Copenhagen

**Workshop #4: NEXT Database**

May 27, 2016, Aarhus University Hospital, Aarhus

**Workshop #5: Status, plan, and deliverables**

November 23, 2016, Aalborg University Hospital, Aalborg

**Workshop #6: Status, plan, and deliverables**

May 19, 2017, Odense University Hospital, Odense

**Workshop #7: Focal points of NEXT Bioinformatics**

November 23, 2017, Rigshospitalet, Copenhagen

**Workshop #8: NEXT Activity report**

May 28, 2018, Aarhus University Hospital, Aarhus

**Workshop #9: NEXT Activity report and perspectives**

November 12–13, 2018, Klitgården, Skagen

Minutes of the meetings can be obtained upon request from the secretariat.



### 3. Common standards for bioinformatics workflows

To ensure consistency of results from high throughput sequencing (HTS) analyses of cancer patients across the country, and thus consistent clinical recommendations, there is a need to develop common standards for bioinformatics processing of raw data. A bioinformatics pipeline consists of many different steps of quality control, alignment, and calibration of the raw data before the final inference of somatic variants can be performed. A variety of different tools for calling somatic variants have been developed and many of these have been compared in terms of the speed and accuracy.<sup>1</sup> Due to the abundance of different tools, building a pipeline is a complex endeavour if all possibilities are considered, so commercial solutions that offer pre-packaged pipelines are sometimes preferred, but many choose to rely on the Genome Analysis Toolkit (GATK) best practices<sup>9</sup>. The GATK is a set of freely available bioinformatics tools with a high level of documentation and support and an accompanying set of best practices for optimal processing of HTS data. It is widely used, notably in the major American cancer sequencing repository, the Genomic Data Commons (GDC) and was already in use by the groups that formed the NEXT Common Standards, so it served as the starting point for the collaboration.

#### **NEXT Common Standards**

To develop the NEXT Common Standards, it was decided to initially focus on the calling of somatic single nucleotide variations and small insertions and deletions based on DNA-seq data, as these are the most widely used targets in HTS-based precision cancer medicine. Local adaptations of the GATK best practices at the participating departments were mapped out, showing that main differences were versions of reference files and genomes, so a common ground based on the GDC HG38 reference genome was agreed upon. Due to some pitfalls in the somatic variant calling performed by the Mutect2 tool in-

a Van Der Auwera et al. (2014) Curr Protoc Bioinformatics 11:1110; <https://software.broadinstitute.org/gatk/best-practices/>





cluded with the GATK, a second variant caller, Varscan2, was added, and a set of custom filters was included to join and filter the output. These recommendations for the common standards and scripts that can be used as building blocks in a bioinformatics pipeline are described in more detail in a report documenting the results of this effort. The participating groups collaborated on this report via the GitHub web-based version control service, and it can be found at: <https://github.com/NEXTBioinformatics/Best-Practices-for-Processing-HTS-Data>

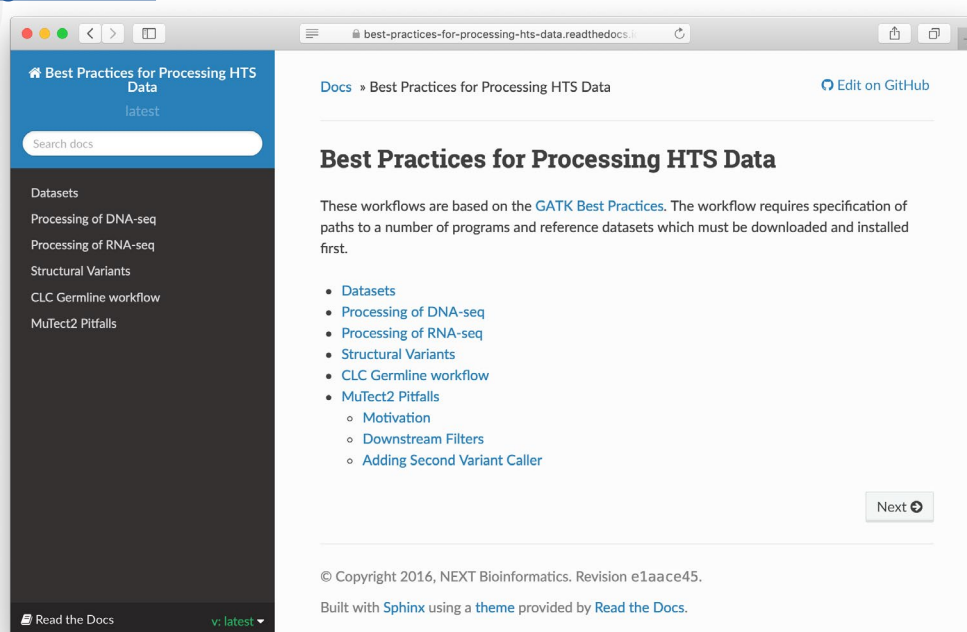


Figure 3.1: Front page of best practices home page.

## Implementation

It was decided that all variants stored in the NEXT variant database, described below, should be processed following the common standards, but since some local pipelines have been accredited for clinical use, they have not been changed. The added variant caller and custom filters are based on experiences from the Department of Molecular Medicine at Aarhus University Hospital (AUH), so their pipeline bears a strong resemblance to the common standards and they have been cited in a work utilizing this approach.<sup>2</sup> The bioinformatics pipeline at the Department of Haematology at Aalborg UH was developed in parallel with the NEXT Common Standards and is based on the principles therein. A scientific paper describing the full workflow is currently



in preparation.<sup>3</sup> The experiences and recommendations from the NEXT common standards may serve as a starting point for building a common Danish bioinformatics pipeline for inferring somatic variants at the Danish National Genome Centre.

## **Future directions**

Although the focus of this working group was somatic variants based on DNA-seq data, there has also been discussion of utilizing RNA-seq data to infer additional information on dysregulated genes in tumours and larger structural variants. These methods are generally less standardized and the participating centres have different aims, but the work was also documented on the GitHub repository. Future work could involve standardizing these workflows across the different centres, and development of methods for detection of over- or underexpressed genes in single samples based on RNA-seq data is also highly relevant. This could be accomplished by forming a reference database of normal expression of genes in different tissues for healthy Danish individuals. Additional information on copy number variation can also be derived from DNA-seq data. This is already done at individual centres and is used for treatment guidance in some instances, but there is no standardization across the country. Finally, there is ongoing work on standardizing workflows for detecting microsatellite instability and tumours with high tumour mutational burden, which will aid in identifying patients who might benefit from immunotherapy.



---

## 4. Biomarker stratification of patients

---

The development of prognostic and predictive biomarkers based on extensive genomic profiling has a great potential in Denmark due to our extensive biobanking facilities and electronic health records. Bioinformaticians contribute to all aspects of this work, from development of new biomarkers for functional studies to implementation of biomarkers for clinical use. This underscores the importance of fostering the education, training, and retention of clinical bioinformaticians to support the Danish potential to be a key player in cancer treatment. NEXT bioinformaticians have participated in all phases of such work over the funding period. Below, we briefly review the work to illustrate the impact of the contributions from NEXT.

### **Sub-classification of cancers by machine learning methods**

NEXT bioinformaticians developed a number of cancer sub-classification tools based on high dimensional molecular signatures using various contemporary machine-learning methods. There has been great interest in classifying the tissue of origin of cancers of unknown primary (CUP), as this may potentially identify possible drug targets. Based on discriminant analysis, RH developed an RNA-classifier based on microarray data.<sup>4</sup> This classifier has been further developed by AUH to include RNA-seq data and a combination of DNA-seq and RNA-seq data, based on restricted multinomial regression techniques,<sup>5</sup> showing that it is possible to identify the original tissue.

Aalborg UH worked with the hypothesis that cancer can be seen as a disease of differentiation by identifying the cell of origin of the cancer for a number of haematological malignancies.<sup>6–10</sup> Their technology is based on restricted multinomial regression techniques with elastic net penalty. They show that the identified origin has prognostic value and treatment predictive potential.



OUH and University of Southern Denmark (SDU) has used archived material and support vector machines to predict the risk of recurrence of low-risk breast cancer and differential diagnosis between thrombocythemia and early prefibrotic myelofibrosis.<sup>11–14</sup> In addition, RH has participated in further characterization of basal-like breast cancer.<sup>15</sup>

## **Systems biology and functional studies**

Biomarker discovery is often conducted by class comparisons of genome-wide profiling of genetic variants and gene and protein expressions. Careful corrections for multiple testing are recommended, and the risk of false positives and spurious correlations is high. Stronger evidence is therefore sought by in silico systems biology studies and in vitro and in vivo functional studies. NEXT bioinformaticians have contributed to a number of systems biology and functional studies over the project period.

NEXT bioinformaticians have been involved in the development of system biological tools for pathway mining and applications, notably within breast cancer and haematological malignancies.<sup>8,9,16–19</sup> NEXT bioinformaticians have also been involved in glioblastoma invasion modelling.<sup>20</sup>

At the Department of Haematology of Aalborg UH, extensive drug screens have been conducted on immortal cell lines originating from haematological cancer patients. These screens have been used to stratify them into sensitive, intermediate, and resistant groups. Differential analyses have been conducted to identify differences in microRNA, gene, and protein expressions between the subgroups. Such differences have been used to build molecular signatures of prognostic significance and potential treatment resistance identification. This work has led to a number of in vitro lentiviral transduction studies to identify the potential of microRNA as a predictor of cytotoxic response potential.<sup>21</sup>

## **Implementation for clinical use**

A clear advantage of a close collaboration between Danish bioinformaticians is that it accelerates the clinical implementation of new algorithms based



on molecular biomarkers. A prominent example is the CUP classification described above, which is now in use in clinical protocols at AUH and RH.<sup>22,23</sup> The cell of origin and resistance classifiers are now being validated in clinical protocols at Aalborg UH. RH has worked to make the pipelines clinically useful in phase 1 studies.<sup>24,25</sup> NEXT Bioinformatics is currently developing a fast platform for implementing new genome-wide tools in clinical applications. Notable examples include NGS-derived tools for calculation of tumour mutational burden and microsatellite instability for use in allocation of immunotherapies. Tools for assessing the clinical relevance of identified somatic variants have been discussed and developed to aid genome interpreters and physicians in their clinical work. Furthermore, RH has studied the potential of following cancer development by ctDNA.<sup>26</sup>



---

## 5. Variant database

---

To search for and recruit patients more efficiently, a platform was built to collect clinical and genomic information about patients at participating centres in Denmark.

In addition to patient recruitment, the platform has three other goals, to provide a good estimate of potential cohort size based on precise inclusion criteria, foster academic research in the field, and act as a prototype for a precision cancer decision support tool.

### Platform

Developed at Aalborg UH, the platform is currently hosted on an Ubuntu 16.04 server at the North Denmark Region's datacentre, which is tier 4-certified. To facilitate its deployment to other hosting solutions, the platform is packaged in a Docker 17.12 container<sup>b</sup>. The platform was developed using Python 3.5<sup>c</sup>, a data-science friendly software language, with a state-of-the-art web framework, Django 2.1<sup>d</sup> and is connected to a Microsoft SQL server 12.0 database. It features access control mechanisms based on roles and centres as well as audit trails. The front-end of the platform was developed in HTML 5, CSS 3 and javascript 1.8 with jQuery 3.2<sup>e</sup> using the Bootstrap 4.1 front-end component library<sup>f</sup>. There was a focus on user experience to make it accessible for people with potentially limited informatics skills.

### Data collection

The principle of the platform is to combine clinical and genomic data and interpretations to make it as relevant as possible for pharmaceutical companies, researchers, and clinicians.

---

<sup>b</sup> <https://www.docker.com/>

<sup>c</sup> <https://www.python.org/>

<sup>d</sup> <https://www.djangoproject.com/>

<sup>e</sup> <https://jquery.com/>

<sup>f</sup> <https://getbootstrap.com/>



## Clinical data

The clinical data are stored in a manner that facilitates visualization of the patient trajectory. Every patient is registered with an anonymized identifier, date of birth, and gender. Each patient belongs to a centre and members of the centre can only see the corresponding data. For every patient, multiple cases can be created, and a case is defined as a primary cancer or relapse. Each case is described by an International Classification of Disease, Version 10 (ICD-10) diagnosis code, a diagnosis date, an International Classification of Disease for Oncology, Version 3.1 (ICD-O-3.1) morphology code, an ICD-O-3.1 topography code, relapse number, treatment regimen, status with a date, and Eastern Cooperative Oncology Group (ECOG) performance status. In the context of a precision cancer medicine research project implemented at Aalborg UH, data are collected using REDCap and utilities were developed to synchronise the data automatically through application programming interface (API) calls.

Patient ID	Sex	Age	Diagnosis	Relapses
1	Male	80	C91.1-B-CCL	5
2	Male	67	C91.1-B-CCL	5
3	Male	70	C90.0-MM	1
9	Male	77	C82.9-FL	2
6	Male	74	C91.1-B-CCL	4
11	Male	87	C82.9-FL	4
14	Male	63	C83.3-DLBC	1
15	Male	36	C81.9-HL	1
16	Male	68	C83.1-MCL	3
19	Female	52	C82.9-FL	1
21	Female	80	C90.0-MM	1
22	Male	76	C83.0-BLL	2
23	Female	73	C90.0-MM	1
24	Male	70	C83.3-DLBC	1
25	Male	76	C91.1-B-CCL	1
26	Female	42	C82.9-FL	1
27	Female	80	C90.0-MM	2
28	Female	70	C82.9-FL	1
30	Female	74	D46.9-MDS	1
32	Male	51	D03.008-LPL	3
33	Female	75	C85.9	2
36	Male	78	C85.9	1
37	Male	70	C82.9-FL	4
38	Female	88	C85.9	1
39	Male	71	C86.9-NTL	1

Figure 5.1: Overview of patients in the platform.

## Genomic data and annotations

For each case, multiple genomic files can be uploaded to the platform. These files are grouped by sample. Samples are defined by an identifier, a sampling date and a tissue type, allowing for storage of information from samples over



multiple locations and monitoring of samples over time. The database currently supports Variant Call Format (VCF) files for Single Nucleotide Variants (SNVs) and insertions-deletions (indels). The Human Genome Variation Society (HGVS) nomenclature is used to uniquely identify variants and is generated by the Variant Effect Predictor (VEP). The VEP also provides annotations such as the allele frequency in the general population and the impact of the variants on the corresponding protein. Future plans include the support of Structural Variations (SVs) including fusions, Copy Number Variations (CNV) and translocations.

Chromosome	Position	Ref > Alt	Frequency	Gene	RSID	Significance	Report	Drugs
01	114,732,907	T > G	115/117	NRAS	127913235	4	2	✓
15	87,860,367	T > G	169/128	NTN3		4	2	-
05	83,526,341	AAAC > A	109/114	VCAN	1485795393	3	1	-
03	108,046,760	T > C	134/132	CD47		2	1	-
03	184,300,457	C > T	111/127	PBM2		2	1	-
17	7,208,789	C > T	26/31	EPN3		2	1	-
01	847,726	C > T	62/68			2	0	-
01	1,000,028	C > T	39/25	HES4	1405048872	2	0	-
01	1,468,939	C > T	22/27	ATA2C		2	0	-
01	15,449,719	G > C	90/90	DD2		2	0	-
01	19,179,129	C > T	112/122	USP4	714960335	2	0	-
01	19,485,458	G > T	36/37	CAF2B		2	0	-
01	20,744,693	A > G	46/51	HP1BP2		2	0	-
01	22,490,582	G > C	123/122	ZBTB45		2	0	-
01	26,125,329	A > G	17/28	POK1L		2	0	-
01	26,781,968	C > T	33/26	ARG1A		2	0	-
01	27,006,437	C > T	235/234	FAM46B	372912903	2	0	-
01	53,210,087	A > G	182/144	CPT2	1488632092	2	0	-
01	76,724,971	G > C	76/83	ACAD4		2	0	-
01	77,063,981	A > C	58/71	ST6GALNAC5		2	0	-
01	113,710,339	G > A	111/126	PHF1		2	0	-
01	119,839,320	G > C	171/177		867360963	2	0	-
01	145,851,164	A > G	159/287			2	0	-

Figure 5.2: Overview of detected variants.





## Interpretation

To advance the understanding of genomic data, efforts have been made to support interpretations generated with Qiagen Clinical Insights (QCI), such as prognosis impacts, targeted treatments, and clinical trials. In addition, we plan to integrate data from the Precision Medicine Knowledge Base (PMKB) and Precision Oncology Knowledge Base (OncoKB) to generate a default set of interpretations for centres that do not use QCI.

The screenshot displays the QCI variant interpretation interface. The main panel shows a variant in the *BRD2* gene (NM\_001199455.1) with a c.966A>T mutation. The variant is located on chromosome 06 at position 32,971,933. The allele frequency is 13.3% (430 reads). The interface includes a 'Somatic mutations' section with buttons for 'Download .VCF', 'Reupload .XML', and 'Delete .XML'. A table below shows the variant details, including chromosome, position, reference/alternate alleles, allele frequency, gene, and HGVS notation. The right panel provides 'Variant annotations from XML', including a table for 'Annotation for gene *BRD2* with transcript NM\_001199455.1' showing reportability (2), significance (2), and actionability (3). Below this, there are sections for 'Prognoses', 'Treatments', and 'Trials', each with a 'No Data' message. The 'Trials' section lists three clinical trials with their phases and titles.

Chr	Pos	Allele frequency	Ref	Alt
06	32,971,933	13.3% of 430	A	T

Reportability	Significance	Actionability
2	2	3

Chr	Position	Ref > Alt	AF	Gene	HGVSc	HGVp
06	32,971,933	A > T	13.3% of 430 reads	<i>BRD2</i>	NM_001199455.1:c.966A>T	-

Showing 1 to 1 of 1 entries (filtered from 1,156 total entries)

Phase	Title
1	Open-Label, Dose Escalation/Expansion Phase Ib Study to Evaluate the Safety, Pharmacokinetics, and Clinical...
1	A Phase 1, Open-label, Dose-finding Study to Assess the Safety, Tolerability, Pharmacokinetics and Prelimin...
1	A Phase I/II Open-label, Dose Escalation Study to Investigate the Safety, Pharmacokinetics, Pharmacodynamic...

Figure 5.3: Example of annotations of variants.



## Variant search

A variant search interface was built to allow for searching for patients across centres based on genomic and clinical data. This functionality can be used to search for patients to be included prospectively or to estimate potential cohort sizes retrospectively for pharmaceutical companies to scale their clinical trials. However, detailed information about the cases is only available to centres that entered data in the platform, giving them tight control of how these data are shared.

429 variants matching your query were found. They are listed below.

### Search criteria

Select EITHER a gene or chromosome to search on (Required)

Gene: TP53 Chromosome: 17

Select a reference genome to use (Required)

Reference Genome: GRCh38

Select a position or range of positions to search at (Optional, click to toggle)

Search for a specific mutation (Optional, click to toggle)

Search on patient data (Optional, click to toggle)

Diagnosis: C91.1 - B-CLL Minimum Age: Gender: Maximum Age:

Search

### Search Results

Copy Excel CSV

Case ID	Patient ID	Project	Diagnosis	Date	Centre	Gene	Chromosome	Position	Mutation	HGVS Code
26	1	AAUH Project	C91.1 - B-CLL	2187-02-01	AAUH	TP53	17	7674243	A > C	NC_000017.11 g.7674243A>C
190	138	AAUH Project	C91.1 - B-CLL	2187-12-07	AAUH	TP53	17	7674221	G > A	NC_000017.11 g.7674221G>A
135	135	AAUH Project	C91.1 - B-CLL	2013-01-10	AAUH	TP53	17	7674243	C > G	NC_000017.11 g.7674243C>G

Figure 5.4: The search interface.



## Precision Cancer Medicine report

By combining interpreted genomic data with clinical data, the platform makes it possible to generate Precision Cancer Medicine (PCM) reports to be used in genomic multidisciplinary team (MDT) evaluations. The idea is to provide external clinicians with an overview of the patient trajectory to allow for a better overall understanding of the patient's situation. A PDF version of the report can be generated on demand for printing and archiving.

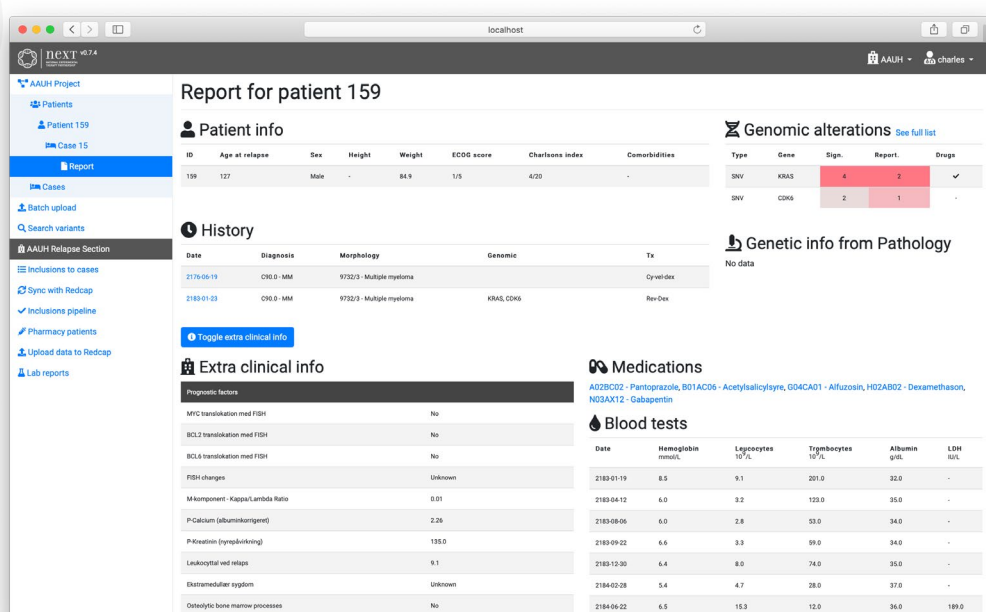


Figure 5.5: Example of tumour board report.



---

## 6. Discussion

---

### **Main results of NEXT Bioinformatics**

Supported by DKK 10 million from the NEXT initiative, NEXT Bioinformatics has facilitated an interactive network of clinical bioinformaticians across Danish university hospitals. The participants of the network have shown a high degree of willingness to share workflows, data, and general knowledge to enable the development, standardization, and implementation of standard workflow. More than 35 people participated in the 10 workshops from the four Danish university hospitals. At the last meeting, the network was expanded to cover the National Genome Center and ELIXIR Denmark. Furthermore, the network has established common bioinformatics standards that are now largely in place at all university hospitals. The group has published more than 20 peer-reviewed papers on biomarker development and a prototype of a searchable database has been developed and deployed in the North Denmark Region's IT infrastructure.

The project has had an impact on Danish health care and the research infrastructure, as the experience from the network build-up was used in the working groups of the newly established Danish National Genome Center. Two networks supported by the Danish Comprehensive Cancer Center (DCCC) plan to use the infrastructures of NEXT Bioinformatics – one focused on somatic variant interpretation and one on establishing a basket trial for treatment of refractory cancer, as well as one 'Knæk Cancer' project on acute myeloid leukaemia.

It was also possible to attract additional funding for NEXT Bioinformatics. For example, DCCC provided DKK 175,000 for consolidating "the national network of clinical bioinformaticians" and AAU's Digital Hub Denmark donated DKK 200,000 of seed money to facilitate the formulation of a larger application for funding the development of decision support tools for personalized medi-



ne based on NEXT Bioinformatics competences. The DCCC network funded a national bioinformatics network workshop on November 12 and 13, 2018 in Skagen and a Broad Institute GATK4 Best Practices workshop on January 29 – February 1, 2019 in Copenhagen.

## **Critical assessment of the work**

A drawback of the work is that the searchable variant database is not yet implemented on a national level and the common standards are not fully implemented or accredited in a clinical setting. However, there are good reasons for this. The decision to develop a database was made by the steering committee during the project period, detracting focus from common standards and biomarker development. The prototype for clinical data collection is not suitable for scaling as it requires much manual work; however, work is in progress to integrate data collection with medicine registries and patient administrative systems.

## **Comparison with other initiatives**

There are other initiatives, but limited information is available. We are not aware of any successful extensive data sharing of somatic variants paired with clinical data. This is probably because other nations' health care systems are based on insurance models, causing big cancer centres with genome medicine to "fight" each other for patients, insurance money, and fame. This starkly contrasts the Scandinavian health care models with one public health care payer. This provides Denmark and the other Scandinavian countries an outstanding opportunity to carry the field forward.

## **Conclusions and suggestions for future work**

Bioinformatics is becoming increasingly important and involved and should therefore be handled by sufficiently large groups with critical mass. Based on the successful achievements of NEXT Bioinformatics, the participating parties have agreed to formally establish a "Danish Society of Clinical Bioinformatics" (DSCB) at the next meeting on May 28, 2019 in Odense. DSCB will be a loosely organized group based on a mailing list and simple regulations. The group will meet once or twice a year and discuss guidelines for newly developed algo-



---

rithms and formulate a funding strategy for national bioinformatics research. The difference between DSCB and other societies (e.g., ‘Dansk Selskab for Molekylær Medicin’ and ‘Dansk Selskab for Klinisk Genetik’) is its clear focus on computational bioinformatics and decision tools for clinical applications. DSCB will seek to engage in the formulation of data support centres for each region.

The individual regions and the National Genome Center will fund standard routine work. However, based on its strong bioinformatics competences, the society will apply for external funding for research in decision support tools for personalized medicine. We see a great development potential in Denmark for studying efficient database construction for variants and patient flows, text mining of electronic health records, new bioinformatics standards for fast data processing, international data sharing, visualization of patient flows, integration with electronic health records, and establishment of clinical evidence.



## 7. Publications

NEXT Bioinformatics has participated in the work of the following publications:

1. Krøigård, A. B., Thomassen, M., Lænkholm, A.-V., Kruse, T. A. & Larsen, M. J. Evaluation of nine somatic variant callers for detection of somatic mutations in exome and targeted deep sequencing data. *PLoS One* 11, e0151664 (2016).
2. Christensen, E. et al. Early detection of metastatic relapse and monitoring of therapeutic efficacy by ultra-deep sequencing of plasma cell-free DNA in patients with urothelial bladder carcinoma. Submitted for publication (2019).
3. Bødker, J. S. et al. Precision medicine workflow in hematological cancers. In preparation (2019).
4. Vikeså, J. et al. Cancers of unknown primary origin (CUP) are characterized by chromosomal instability (CIN) compared to metastasis of known origin. *BMC Cancer* 15, 151 (2015).
5. Søndergaard, D., Nielsen, S., Pedersen, C. N. S. & Besenbacher, S. Prediction of primary tumors in cancers of unknown primary. *J. Integr. Bioinform.* 14 (2017).
6. Falgreen, S. et al. hemaClass.org: Online one-by-one microarray normalization and classification of hematological cancers for precision medicine. *PLoS One* 11, e0163711 (2016).
7. Nørgaard, C. H. et al. Subtype assignment of CLL based on B-cell subset associated gene signatures from normal bone marrow – A proof of concept study. *PLoS One* 13, e0193249 (2018).
8. Schönherz, A. A. et al. Normal myeloid-progenitor-cell-subset gene signatures for acute myeloid leukemia are associated with prognosis. Submitted for publication (2018).
9. Bødker, J. S. et al. A multiple myeloma classification system that associates normal B-cell subset phenotypes with prognosis. *Blood Adv.* 2, 2400–2411 (2018).
10. Vesteghem, C. et al. Recommendations for the implementation of the FAIR Data Principles in Personalized Cancer Medicine. In preparation (2019).
11. Sørensen, K. P. et al. Long non-coding RNA expression profiles predict metastasis in lymph node-negative breast cancer independently of traditional prognostic markers. *Breast Cancer Res.* 17, 55 (2015).
12. Block, I. et al. Differential microRNA expression predicts recurrence-free



- survival in systemically untreated breast cancer. In preperation (2019).
13. Block, I. et al. Association of miR-548c-5p, miR-7-5p, miR-210-3p, miR-128-3p with recurrence in systemically untreated breast cancer. *Oncotarget* 9, 9030–9042 (2018).
  14. Skov, V. et al. A 7-gene signature depicts the biochemical profile of early prefibrotic myelofibrosis. *PLoS One* 11, e0161570 (2016).
  15. Kinalis, S., Nielsen, F. C., Talman, M.-L., Ejlersen, B. & Rossing, M. Characterization of basal-like subtype in a Danish consecutive primary breast cancer cohort. *Acta Oncol.* 57, 51–57 (2018).
  16. Alcaraz, N. et al. Robust de novo pathway enrichment with KeyPathwayMiner 5. *F1000Research* 5, 1531 (2016).
  17. List, M. et al. KeyPathwayMinerWeb: online multi-omics network enrichment. *Nucleic Acids Res.* 44, W98–W104 (2016).
  18. Alcaraz, N. et al. De novo pathway-based biomarker identification. *Nucleic Acids Res.* 45, e151–e151 (2017).
  19. Batra, R. et al. On the performance of de novo pathway enrichment. *Syst. Biol. Appl.* 3 (2017).
  20. Jensen, S. S. et al. Establishment and characterization of a tumor stem cell-based glioblastoma invasion model. *PLoS One* 11, e0159746 (2016).
  21. Marques, S. C. et al. High miR-34a expression improves response to doxorubicin in diffuse large B-cell lymphoma. *Exp. Hematol.* 44, 238–46.e2 (2016).
  22. Overby, A., Duval, L., Ladekarl, M., Laursen, B. E. & Donskov, F. Carcinoma of unknown primary site (CUP) with metastatic renal-cell carcinoma (mRCC) histologic and immunohistochemical characteristics (CUP-mRCC): results from consecutive patients treated with targeted therapy and review of literature. *Clin. Genitourin. Cancer* 17, e32–e37 (2018).
  23. Gravholt, C. H. et al. A rare case of embryonal carcinoma in a patient with Turner Syndrome without Y chromosomal material but mutations in KIT, AKT1, and ZNF358 demonstrated using exome sequencing. *Sex Dev.* 11, 262–268 (2017).
  24. Tuxen, I. V. et al. Personalized oncology: genomic screening in phase 1. *APMIS* 122, 723–733 (2014).
  25. Tuxen, I. V. et al. Copenhagen Prospective Personalized Oncology (CoPPO) - Clinical utility of using molecular profiling to select patients to phase 1 trials. *Clin. Cancer Res.* In press (2018).
  26. Ahlborn, L. B. et al. Circulating tumor DNA as a marker of treatment response in BRAF V600E mutated non-melanoma solid tumors. *Oncotarget* 9, 32570–32579 (2018).





---

## 8. Acknowledgements

---

We thank Hans Erik Johnsen for his strong support in establishing and supporting the development of NEXT Bioinformatics.



---

## 9. Participants

---

### **Rigshospitalet/Copenhagen University**

Christina Westmose Yde  
Savvas Kinalis  
Ane Yde Schmidt  
Lars Jønson  
Ole Winther  
Filipe Garrett Vieira

### **Herlev University Hospital/Copenhagen University**

Estrid Høgdall  
Tim Svenstrup Poulsen

### **Odense University Hospital/University of Southern Denmark**

Mark Burton  
Henrik Ditzel  
Torben Kruse  
Mads Thomassen  
Martin Larsen  
Jan Baumbach  
Fabio Vandin  
Lars Andersen  
Nicolas Alcaraz

### **Aarhus University Hospital/Aarhus University**

Torben Ørntoft  
Søren Besenbacher  
Maria Bach Laursen  
Britt Elmedal Laursen  
Christine Kassentoft  
Michael Knudsen  
Jan Larsen

### **Aalborg University Hospital/Aalborg University**

Martin Bøgsted



---

Karen Dybkær  
Hans Erik Johnsen  
Rasmus Froberg Brøndum  
Charles Vesteghem  
Anders Sune Petersen  
Mads Sønderkær  
Lasse Hjort Jakobsen  
Anna Schönherz  
Anders Ellern Bilgrau  
Julie Støve Bødker  
Jonas Have  
Kasper Lindblad Nielsen  
Dennis de Weerd

### **External delegates**

Claus Lindbjerg Andersen, Aarhus University Hospital/Aarhus University  
Rune Møllegaard Friborg, Aarhus University  
Søren Brunak, Copenhagen University  
Cathrine Jespersgaard, The National Genome Center  
Marco Zingales, The National Genome Center  
Britta Smedegaard Andersen, NEXT  
Ask Aabenhus, NEXT



# NEXT Bioinformatics

NEXT Bioinformatics  
<https://nextpartnership.dk/>

February 2019